MULTIPLE SCLEROSIS JOURNAL | MSJ

Meeting Review

Improving the efficiency of clinical trials in multiple sclerosis

Multiple Sclerosis Journal
2023, Vol. 29(9) 1136–1148
DOI: 10.1177/13524585231189671

Ruth Ann Marrie, Maria Pia Sormani, Sean Apap Mangion, Francesca Bovis, Winson Y Cheung, Gary R Cutter, Peter Feys, Michael D Hill, Marcus Werner Koch, Morgan McCreary, Ellen M Mowry, Jay JH Park, Fredrik Piehl, Amber Salter and Jeremy Chataway; on behalf of the participants in the International Conference on Innovations in Clinical Trial Design & Enhancing Inclusivity of Clinical Trial Populations

Abstract

Background: Phase 3 clinical trials for disease-modifying therapies in relapsing-remitting multiple sclerosis (RRMS) have utilized a limited number of conventional designs with a high degree of success. However, these designs limit the types of questions that can be addressed, and the time and cost required. Moreover, trials involving people with progressive multiple sclerosis (MS) have been less successful.

Objective: The objective of this paper is to discuss complex innovative trial designs, intermediate and composite outcomes and to improve the efficiency of trial design in MS and broaden questions that can be addressed, particularly as applied to progressive MS.

Methods: We held an international workshop with experts in clinical trial design.

Results: Recommendations include increasing the use of complex innovative designs, developing biomarkers to enrich progressive MS trial populations, prioritize intermediate outcomes for further development that target therapeutic mechanisms of action other than peripherally mediated inflammation, investigate acceptability to people with MS of data linkage for studying long-term outcomes of clinical trials, use Bayesian designs to potentially reduce sample sizes required for pediatric trials, and provide sustained funding for platform trials and registries that can support pragmatic trials.

Conclusion: Novel trial designs and further development of intermediate outcomes may improve clinical trial efficiency in MS and address novel therapeutic questions.

Keywords: Multiple sclerosis, clinical trials, platform trials, adaptive trial designs, Bayesian statistics, futility trials

Date received: 8 March 2023; revised: 1 June 2023; accepted: 12 June 2023

Correspondence to:
RA Marrie
Departments of Internal Medicine and Community Health Sciences, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, GF532-820 Sherbrook Street, Winnipeg, MB R3A 1R9, Canada.
rmarrie@hsc.mb.ca

Ruth Ann Marrie
Departments of Internal Medicine and Community Health Sciences, Max Rady College of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada

Maria Pia Sormani
Department of Health Sciences, University of Genoa, Genoa, Italy/IRCCS Ospedale Policlinico San Martino, Genoa, Italy

Sean Apap Mangion
Queen Square Multiple Sclerosis Centre, Department of Neuroinflammation, UCL Queen Square Institute of Neurology, Faculty of Brain Sciences, University College London, London, UK

Francesca Bovis
Department of Health Sciences, University of Genoa, Genoa, Italy

Winson Y Cheung
Department of Oncology, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Gary R Cutter
Department of Biostatistics, University of Alabama at Birmingham, Birmingham, AL, USA

Peter Feys
REVAL Rehabilitation Research Center, REVAL, Faculty of Rehabilitation Sciences, Hasselt University, Hasselt, Belgium/Universitair MS Centrum, UMSC, Hasselt, Belgium

Introduction

Phase 3 clinical trials for disease-modifying therapies (DMTs) in relapsing-remitting multiple sclerosis (RRMS) have evolved from a limited number of Phase 2/3 designs and have led to successful development of multiple drugs used as monotherapies. However, clinical trials for progressive multiple sclerosis (MS) have been less successful, as have trials of rehabilitation therapies.[1,2] Traditional Phase 3 designs are not well-suited to test complex treatment strategies. Moreover, they have constraints related to the types of questions that can be addressed, who participates, and the time and cost required. Therefore, alternatives such as platform and adaptive designs with appropriate intermediate and primary outcomes may be more efficient and appropriate depending on the question being asked and the population being considered. The increasing adoption of electronic health records that may support registry-based trials, acceptance of more pragmatically designed trials, and acceptance of Bayesian designs also offers the opportunity to accelerate testing of therapies for progressive MS and to address novel questions. In December 2022, an international group of investigators in MS, epidemiology, biostatistics, rehabilitation, and clinical trial design met under the auspices of the International Advisory Committee on Clinical Trials in MS, sponsored by the European Committee on Treatment and Research in MS and the US National MS Society (see Supplemental Appendix 1). Based

on that meeting, we review novel designs and the role of intermediate and composite outcomes as mechanisms to improve clinical trial design efficiency and to answer new questions. The use of real-world data in observational study designs to evaluate treatment effectiveness is reviewed elsewhere.[3,4]

Novel designs

Clinical trials exist on a continuum between explanatory (i.e. efficacy—does the intervention work under ideal conditions) and pragmatic (i.e. effectiveness—does the intervention work under usual conditions). The Pragmatic–Explanatory Continuum Indicator Summary 2 (PRECIS-2) wheel allows investigators to consider and describe how pragmatic or explanatory their trial will be, considering multiple elements.[5] These include participant eligibility and recruitment, setting, organization (for delivery of the intervention), degree of flexibility with respect to delivery and adherence by participants, follow-up, the primary outcome, and the primary analysis. For example, a pragmatic trial may be embedded within a registry which is used as the mechanism for recruitment, data collection, and follow-up (COMBAT-MS Trial (NCT03193866), HOPE-Covid19 Supplemental Appendix Box e1).[6] Most trials testing interventions in MS to date fall on the more explanatory end of the spectrum. These trials employ strict inclusion and exclusion criteria to minimize heterogeneity in the population,[7] randomize participants, and rigorously control delivery of the intervention and its assessment. Regulators consider these criteria when labeling drugs for use,[8] yet these trials may not adequately inform practice because the populations enrolled may not reflect those seen in clinical practice, the investigators may be more experienced and conduct more frequent assessments thereby enabling earlier detection and mitigation of harms, and adherence to the intervention may be higher. These issues are not unique to studies of pharmacologic interventions. Consequently, benefits of the intervention may be overestimated and harm underestimated. These designs are also poorly suited to testing complex treatment strategies, such as whether initiation of higher efficacy versus lower efficacy DMT after diagnosis of MS produces better long-term clinical outcomes without increased harm, treatment de-escalation, or (complex) rehabilitation interventions.[9,10] Also, many clinical trials in MS have used conventional trial designs with pre-specified treatment arms, fixed sample size (or number of events), and one primary final analysis. This has contributed to high costs and the slow pace of drug development, which is a particular concern for progressive MS, where a few approved therapies exist.

Pragmatic and novel trial designs seek to address some of the limitations of conventional clinical trials (see Table 1 for key features, strengths, limitations, and potential applications). Pragmatic trials may cost less for the investigator to conduct while retaining the benefits of randomization, but larger sample sizes may be required due to smaller effect sizes. This may be due in part to greater heterogeneity of the sample recruited, and less consistent or controlled data collection methods; data quality also can be a concern.[6,11] Pragmatic cluster randomized designs are useful for testing health system–level interventions such as changes in care pathways.[12] Clusters, such as specific clinics or hospitals, are randomized to an intervention, but outcomes are measured at the patient level. Because patient-level data within a cluster are correlated (intra-class correlation), analyses must account for the correlations; sample sizes using clusters as the unit of randomization are often larger than conventional trials, and power is achieved by increasing the number of clusters, not participants. The stepped wedge design is a variant of the cluster randomized design,[13] in which the cluster is randomized to the timing of the intervention, which is implemented in a staggered fashion. Typically, all clusters ultimately receive the intervention; thus, there are also features of a crossover design and avoidance of ethical concern when testing a hypothesis related to an intervention that is very likely to be beneficial. Herein, we highlight complex innovative designs, designs for combination therapies, and benefits of data linkage.

Complex innovative designs

Clinical trials with adaptive trial designs refer to a group of clinical trial designs that offer pre-planned opportunities to use accruing data to modify aspects of an ongoing trial.[14] They use a pre-specified statistical analysis plan to preserve the validity and integrity of the trial, such as control of type I error rates. Modifiable components include the eligibility criteria (adaptive enrichment design), sample size, allocation ratio (response-adaptive randomization), study intervention, or dose.[14,15] Sequential designs, the most common type of adaptive trial design utilized, can reduce the required sample size and shorten trial duration by allowing the trial to be stopped early for superiority or futility in the case of overwhelming evidence. However, the potential efficiency gained from using adaptive trial designs comes with statistical and operational complexities.[16]

The slow process of trial development and start-up of trials contributes to slow progress in identifying efficacious new therapies. Platform (or multi-arm

Michael D Hill
Departments of Clinical Neurosciences, Community Health Sciences, Medicine, and Radiology, Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Marcus Werner Koch
Departments of Clinical Neurosciences, Community Health Sciences, Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Morgan McCreary
Amber Salter
Department of Neurology, Section on Statistical Planning and Analysis, UT Southwestern Medical Center, Dallas, TX, USA

Ellen M Mowry
Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

Jay JH Park
Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada

Fredrik Piehl
Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

Jeremy Chataway
Queen Square Multiple Sclerosis Centre, Department of Neuroinflammation, UCL Queen Square Institute of Neurology, Faculty of Brain Sciences, University College London, London, UK/National Institute for Health Research, University College London Hospitals, Biomedical Research Centre, London, UK/ Medical Research Council Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, University College London, London, UK

Table 1. Key features, strengths, limitations, and applications of novel trial designs in multiple sclerosis.

Design	Key features	Strengths	Limitations	Appropriate applications
Adaptive trials	• Various possible types and designs; ultimate advantage/ focus being the inclusion of defined points/criteria for modification of the trial design to achieve greater efficiency	• Savings for time, cost, sample size, patient exposure.	• Greater statistical and logistical requirements.	• Broad; depends on goal and type of trial utilized *The full table in Supplemental Materials outlines these in depth*
Platform multi-arm multi-stage	• Multiple therapies evaluated on a single-disease group, divided into arms, against a single control group. • Predefined interim analyses determine whether a treatment will proceed to next stage analysis or discontinued. • Additional therapies can be added at predefined time points.	• Built-in progression from Phases 2 to 3 if interim outcomes are met. • Opportunity to increase the control group rather than experimental arms (potential to limit cost). • Opportunity to test new hypotheses/ arms during recruitment, while still controlling family-wise errors. • Shorter time and cost requirements compared to individual Phase 2/3 trials for different agents.	• Inherent operational and design challenges. • Complicated design/statistical issues depending on arm retention/additions, with implications for funding and logistics. • Underlying assumption is that all treatments work equally well under the null hypothesis (i.e. no one is better than any other). • Greater upfront cost. • Rely on valid, reliable intermediate outcomes that accurately predict the primary outcome.	• For use when multiple promising treatments for Phase 2/3 studies are available, with no strong belief that one treatment will be more effective than another. • Requires availability of adequate funding, and number of patients for enrollment. • Requires suitable intermediate outcome measure/s which correlates with the primary outcome measure (when the platform is designed for early phase adaptive trials).
Futility designs	• Phase 1/2 screening trial design for treatments of interest. • The null hypothesis is that the treatment of interest will increase the number of treatment successes by a minimal clinically significant amount.	• Optimizes early phase trial times. • Requires minimal sample sizes, utilizing historic controls as the trial's control arm and to generate the likely outcome without treatment effect, and the clinically significant effect.	• Requires accurate predictions of likely natural disease progress without effective treatment, and agreement that the proposed treatment success rate is indeed clinically significant. • Risk of bias from unblinded treatment and reliance on historic controls. • Positive trial result does not support treatment efficacy but only indicates non-futility.	• For use in Phase 1 and 2 studies screening translational treatments of interest rapidly. • Particularly suitable for repurposed drugs.
Pragmatic cluster randomized	• Randomization, or control of exposure, to new treatment/s to clusters of patients rather than individuals. • Utilization of established registries to monitor outcomes.	• Benefit of limiting three major unquantifiable errors in large-scale interventions, time-dependent confounding, the Hawthorne effect, and regression-to-the-mean. • Significant cost-saving by utilizing registries for monitoring outcomes.	• Sample size estimations and statistics are complicated and need to consider the number of clusters, calendar-time, inter-cluster correlations, observations per cluster, and ultimate detail of the design. • Greater potential for underlying variability of quality of data collection, and missing data fields.	• Suitable for large-scale projects, such as national or healthcare-wide interventions, assessment of clinical pathways, or initiation of electronic records.

Note: Refer to full table in Supplemental Material for details of the different trial designs within each heading.

multi-stage, MAMS) protocols are characterized by planned flexibility with respect to interventions (Supplemental Appendix Figure e1), but remain infrequent designs in neurology.[17,18] They enable several interventions to be evaluated using a common control that can change over time by establishing an overarching protocol document referred to as a master or core protocol.[17] Platform trials aim to act as a long-term or perpetual clinical trial infrastructure in which different interventions can enter and leave at different times.[18] Platform trials are often conducted with adaptive trial designs to screen multiple interventions rapidly.[19] For instance, the STAMPEDE (Systematic Therapy in Advancing or Metastatic Prostate Cancer: Evaluation of Drug Efficacy) trial is the longest ongoing adaptive platform trial and changed the standard of prostate cancer care.[20,21] For progressive MS, the OCTOPUS trial (ISRCTN 14048364) is currently testing metformin and alpha-lipoic acid with plans to add additional therapies.

Platform trials require more time and financial investment to set-up initially than conventional trials, but the use of a common infrastructure ultimately reduces the costs and time required relative to number of interventions assessed.[22] These trials require careful planning, collaboration, and input with respect to clinical, operational factors, and statistical considerations.[23] Appropriate outcome measures are critical. Interim outcomes, such as imaging outcomes commonly used as intermediates in Phase 2 designs in RRMS, can be used to rapidly identify unsuccessful arms. However, the intermediate outcomes would need to be adapted to the study population, question of interest, and therapy's mechanism of action; for example, ocular coherence tomography in a neuroprotection trial. These intermediate outcomes must be able to accurately predict the effect of the intervention on the primary outcome of interest.

Other designs that are conducted under the master protocol framework include basket (single treatment applied to multiple conditions) and umbrella trials (single condition treated with multiple approaches),[17] which originated from precision oncology, driven by the increased power of genomic differentiation of cancer subtypes.[24] These designs are not yet applicable to MS.

The futility, or non-superiority, design can be employed to rapidly screen potential therapies, particularly when considering repurposed drugs.[25] First used in oncology, they are starting to be applied in MS.[26] Futility studies do not use contemporary controls but rely on a clear understanding of the natural history of the disease, that is, how the disease would evolve in untreated individuals, and whether this historical rate is still applicable. The simplest design employs an open-label design with one arm. The failure threshold is based on the natural history of disease progression, and the success threshold is based on existing treatments or other clinically meaningful effect. The Gehan[27] model uses two stages. In the first stage, participants are enrolled, and if there are no treatment successes, the intervention is deemed futile. If there are $\geqslant 1$ treatment successes, the trial moves on to the next stage, adding more participants based on the number of successes in the first stage. This design is challenging due to the variability in sample size and permits ineffective drugs to move forward by allowing a drug with only one treatment success in stage one to do so. Fleming extended this design to include two or three stages, to include a fixed rather than variable sample size, and to allow termination after the first stage if the treatment is very effective or very ineffective.[28] The popular Simon two-stage model is a further development of these earlier futility designs. It comes in two design variations that affect sample size and futility thresholds.[29] The optimal design minimizes the expected sample size for the first stage, while the minimax minimizes the overall sample size. Futility trials allow for much smaller sample sizes than traditional randomized controlled designs, but they have a higher risk of bias due to reliance on historical controls and the lack of blinding. Futility designs should not be used in populations where spontaneous improvement occurs, and the rate of disease worsening must be predictable; notably, progressive MS may not always meet these conditions. Futility trials can only identify interventions which merit further testing, but are not definitive as they do not provide support of efficacy, only lack of futility. To date the futility design has been used to test domperidone in secondary progressive MS and hydroxychloroquine in primary progressive MS (Box 1).[26,30]

Bayesian designs

Bayesian trial designs have become increasingly used, particularly in oncology and early phases of drug and biologics development,[31] and medical device trials.[32] The US Food and Drug Administration has also supported development of Bayesian adaptive trials.[31] In the classic frequentist framework, the *p*-value estimated in clinical trials represents the probability of obtaining the observed or a more extreme treatment effect, given that the null hypothesis is true. The Bayesian methodology allows us to estimate the probability of each size of a treatment effect, after observing

Box 1. Planning a Simon two-stage (minimax) design study for primary progressive MS.[22]

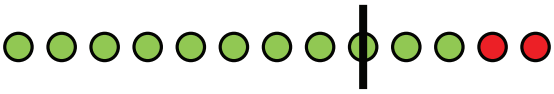
Natural history: 40% of people with primary progressive MS fulfilling the inclusion criteria are expected to worsen per year*

Clinically meaningful benefit: 20% of trial participants worsen per year

Set threshold based on: type 1 error: 5%, type 2 error: 20%

Hydroxychloroquine (HCQ) successful if < 10 of 35 participants worsen ⩾20% on timed 25-foot walk

Stage 1 (interim) analysis
$n = 13$ participants, max 4 can worsen for the trial to continue



(In the HCQ in PPMS trial, 2 of 13 participants worsened, so the trial continued into the second stage)

If successful, proceed to next stage. Enroll 22 more participants (total $n = 35$)

Stage 2 (final) analysis
$n = 35$, max 9 can worsen for the drug to be deemed non-futile

(In the HCQ in PPMS trial, 8 of 35 participants worsened, so HCQ was deemed non-futile)
*Based on the INFORMS and PROMISE trials

the results of the trial (the so-called posterior probability distribution). Two components are needed to estimate a Bayesian posterior distribution: the probability of the observed data given the value of a parameter (i.e. the frequentist probability coming from the experiment) and the prior knowledge about the parameter (the prior distribution).

The Bayesian approaches are usually adopted for two reasons. The first is to overcome mathematical limitations in the frequentist framework or to have an independent confirmation of treatment results. In this case, the prior distribution is often defined to be "non-informative." That is, a distribution that assigns all the values the same probability and suggests no prior knowledge about the treatment effect. The second is to "update" prior evidence such as the size of a treatment effect with the results of our experiment. In this case, we use an "informative" prior, that is a probability distribution telling us the prior evidence about the effect size of a drug. The main obstacle to the application of Bayesian methods is the subjectivity of the prior distribution, but this can be built using evidence-based methods, such as meta-analytic predictive priors that are based on a meta-analysis of previous studies.[33,34]

The Bayesian clinical trials often have better operating characteristics, such as lower sample size requirements, when significant prior information is available,[35] than clinical trials conducted under frequentist statistical framework, while being able to meet the frequentist type I error rate control.[36] A possible application in MS research of Bayesian approaches is in trials for children with MS testing therapies that have been evaluated in adults (see Box 2 for an example).[37,38] Conventional frequentist clinical trials for pediatrics may require similar sample size requirements as adult trials. Since MS is rare in children, it is often not feasible to conduct clinical trials with large sample sizes unless the trial is be powered based on large, potentially unrealistic, treatment effects. The pediatric MS trials are likely too small to detect clinically meaningful effects as a result. Under the Bayesian framework, we may explicitly apply adult clinical trial data as the prior evidence to be incorporated with the pediatric data (the likelihood) to overcome this challenge of conducting pediatric MS trials. If a drug candidate is shown to be efficacious in adults, it might be plausible that the same drug may also be efficacious in children with MS.

Designs for combination therapies

Most clinical trials of DMTs in MS have tested single (mono) therapies.[39,40] Improvements in the management of the disease may require developing combination therapies that address multiple pathophysiologic mechanisms underlying disease activity and progression in MS.[41] Broadly, combination therapies use two or more therapies together, either simultaneously, as add-on therapy, or as sequenced approaches.[40] They may share a common target or pathway or may target

Box 2. Bayesian trials for pediatric multiple sclerosis.[37,38]

Frequentist approach	Bayesian approach
Inferences based on all possible data generated by the experiment, but all possible data not actually observed	Dependent upon data observed in the current experiment
Parameters fixed, but true parameter unknown	Parameters represented by probability distribution
Point estimates include maximum likelihood or least squares estimates	Point estimates include summary statistics of the posterior distribution (e.g. mean, median, and mode)
p-value represents probability of observing the same results, or more extreme results in the sample, if the null hypothesis is true in the population from which the sample is drawn	Posterior probability represents the probability of the null hypothesis
X% confidence interval represents an interval that would contain the true parameter value in X% of repeated samples	X% credible interval represents an interval that contains the true parameter value with X% probability
Safety and Efficacy of Teriflunomide vs Placebo in Paediatric Multiple Sclerosis (TERIKIDS) Study	
Outcome: time to first relapse Frequentist approach (as published): ↓ HR = 0.66; 95% CI = 0.39–1.11 →	Bayesian approach: incorporate information from pooled effects of the TEMSO and TOWER trials in adults (HR = 0.68; 95%CI = 0.58–0.79) ↓ HR = 0.67; 95% CI = 0.51–0.87

different disease mechanisms. Such combination therapies may have additive or synergistic effects, thereby improving treatment response or mitigate harm by allowing the use of lower doses of the individual therapies. Challenges in identifying successful combinations of therapies in MS include the poor ability of preclinical models to predict additivity or synergism, potential interference of one therapy with the other's efficacy, and unintended off-target effects due to the combination. Choice of outcome measure is more complicated if multiple mechanisms are targeted, and implementation of the trial is more complex.

Modern model-based early phase trial designs could be utilized to improve the efficiency and potential success of identifying successful combinations of therapies in the future.[42] Later phase trials using factorial designs or adaptive trial designs (e.g. multi-stage multi-arm) can be considered in evaluating efficacy of therapeutic combinations.[43] Regulators usually require that a combination therapy is compared to each component separately to demonstrate the combination is better than either alone to accept the risk of multiple drugs when one might suffice. Factorial designs test the effect of two or more therapies with multiple levels that are crossed. These designs require that each intervention can be administered without changing the dose when administered in combination with the other. Factorial designs allow more efficient testing of two interventions, as fewer patients are needed than if the therapies are tested separately, although interactions between therapies can increase sample size. The COGEx trial, for example, is a randomized, blinded, sham-controlled trial that is testing whether the combination of cognitive rehabilitation and exercise interventions is more effective than the individual therapies and control conditions for improving processing speed deficits in people with progressive MS.[44]

Data linkage

In observational studies, linkages are often used to obtain data that are not available in a single source to support data quality assessments and to expand the scope of research inquiries. In contrast, a scoping review of publications between 1945 and 2016 identified only 113 clinical trials that used linkage,[45] including one in MS.[46,47] Yet, extending clinical trials across the explanatory-pragmatic continuum by linkage to external data sources, such as health claims data (Box e1), cancer or vital statistics registries offers the opportunity to examine a broader range of outcomes during the trial or to evaluate long-term effectiveness and safety of pharmacologic and/or non-pharmacologic interventions (e.g. extension phase) while minimizing the burden on participants and clinicians.[48] Notably, 18.6% of the linkage-based studies in the scoping review identified long-term benefits, and 9% identified harms that were shown only in the extension phase.[45] The use of real-world data sources can support examination of outcomes relevant to the health system and society such as hospitalizations, admission to long-term care, healthcare costs, or use of social services.[49] This is important when clinical trials have

Table 2. Considerations related to use of data linkage in clinical trials.

Advantages	Disadvantages	Other considerations
• Low burden on participants and trial sites	• Endpoints limited to those captured in the data source. Complex clinical, performance-based, and imaging measures often not captured	• Participant consent
• Comparatively low cost to primary data collection	• Endpoints need to be validated and may not be available across all jurisdictions of interest	• Nature of linkage (probabilistic or deterministic)
• May minimize selection bias and loss to follow-up if all participants in prior trial agree to participate[51]	• Healthcare use may differ due to participant characteristics unrelated to the intervention	• Completeness and accuracy • Timing of extensive phase relative to parent trial
• Can establish external comparator group[52]		• Availability of standardized definitions for variables to be used[53]
• Ability to capture outcomes non-traditional outcomes such as admission to long-term care and use of social services		• Time and administrative burden of data access and regional variation in access processes

short follow-up of 2–3 years, but where long-term outcomes are relevant for clinical decisions and policies. Extension trials in MS typically use a traditional site-based approach but can be designed to rely exclusively on data linkages across multiple jurisdictions.[49] Alternatively, hybrid models that capture typical clinical trial outcomes can be enriched by data linkages to capture additional outcomes. The decision to use data linkages should consider the reason for the linkage, desired endpoints, timing of the data linkage relative to the parent trial, participant burden, costs, data quality, and response of regulators (Table 2).[49] A feasibility study in a lymphoma population suggests that such extension studies are highly acceptable to most clinical trial participants and research ethics boards.[50]

Outcome measures

A critical component of any clinical trial is the outcome measures; the primary outcome determines trial duration and sample size. Clinical trials in MS may include one or more types of outcome measures, including clinician-assessed, performance-based, patient-reported outcomes, and biomarkers. The primary outcome measure used has varied across trial phases and populations in MS. Phase 2 trials in RRMS generally rely on intermediate outcomes such as magnetic resonance imaging (MRI) endpoints, whereas Phase 3 trials have relied on clinical endpoints such as relapse rates or disability progression (the Expanded Disability Status Scale [EDSS] score). Similarly, most Phase 3 clinical trials in progressive MS have relied on clinician-assessed progression as the primary outcome,[54] despite poor responsiveness of the EDSS to evolution of progressive disease, particularly among individuals with higher baseline EDSS scores. Although preservation of cognition is rated as important by people with MS, it is insufficiently assessed in most DMT trials,[54] often relying on a single performance–based measure. Responsiveness of some, although not all, commonly used patient-reported outcomes is also inadequate or poorly understood,[55,56] and clinical trials frequently do not include outcomes that reflect the impact of interventions on activities and participation for people with MS. Outcome measures need to be meaningful, reflect the underlying construct of interest, be valid, reliable, and responsive.

Composite outcomes

Composite endpoints combine multiple clinical outcomes (components) that reflect a common underlying disease process. The endpoint is defined as the occurrence of any of the specified outcomes in a patient. For example, evidence of disease activity is a composite endpoint created using additive meaningful components that includes relapses and changes in disability and MRI activity.[57] Another example would be disability progression identified based on meaningful worsening of the EDSS (0.5–1.0 point depending on baseline value), the timed 25-foot walk (20% worsening),[58] the nine-hole peg test (20% worsening),[59] or the Symbol Digit Modalities Test (4 point decrease in raw score).[60] Composite endpoints are distinct from composite indicators which are separate items that are combined to create a new summary variable.[61] Variables that are continuous can be combined through the use of a simple or weighted average. The Multiple

Sclerosis Functional Composite is a composite outcome that includes a measure of lower limb function, upper limb function, and cognition,[62] by averaging *z*-scores of the original variables. A simple average works well when the strength of the relationship between the components and the independent variables is similar. Weighted averages assign weights derived from principal components analysis or from prior studies to the component variables that comprise the composite. Composite endpoints offer the opportunity to more fully capture the multi-dimensional experience of MS. The use of composite outcomes can increase event rates, thereby increasing statistical power, enabling shorter clinical trials with smaller sample sizes. However, the observed treatment effect may not apply to all components of the composite, thereby complicating interpretation and reporting with respect to individual components.[63] Moreover, if one component is not affected by the treatment, statistical power to detect an effect on the composite may be reduced;[64] power decreases as the number of non-responsive components increases. This issue becomes more complicated if responsiveness of the components varies with the baseline status (e.g. overall disability level) of the participant. Furthermore, if treatment effects on components are in opposite directions, interpreting study findings is complex.

Intermediate outcomes

Clinical trials that employ intermediate outcomes require smaller sample sizes, are shorter in duration, and have higher statistical power. A clinical outcome is a direct measure of how the patient feels, performs, or survives. An intermediate outcome is a measure of a function or symptom (such as pain) which is not the ultimate endpoint of disease. Intermediate outcomes include biomarkers and replacement endpoints that are considered to assess the causal pathway through which the intervention affects the true outcome of interest;[65] the latter should be meaningful to the patient. In rehabilitation clinical trials, there are similar considerations with respect to outcome measures. In rehabilitation, a commonly used conceptual framework is that of the International Classification of Functioning, Disability and Health Framework. This framework includes the core domains of body function and structures, activities, participation, personal, and contextual factors. In this framework, we could consider the intermediate outcomes as those related to the "body function" under study such as lower limb strength or muscle tone. The key clinical outcomes would be those related to activities as could be measured using a six-minute walk,[66] and participation, as could be measured through social participation or community life, health-related quality of life, or functional independence in daily activities (see companion paper for further discussion of these issues in rehabilitation[67]). Surrogate outcomes are intermediate outcomes that meet specific criteria. Specifically, the intermediate outcome must be strongly associated with the outcome of interest, and the effects of the intervention on the outcome of interest are fully captured via the intermediate.

Intermediate outcomes can be used for several reasons.[68] They can be used as endpoints in Phase 2 clinical trials or at the interim stage of a platform trial to determine whether the intervention merits further evaluation in a longer, more costly Phase 3 trial. In Phase 3 trials, they can be used as surrogates to allow decisions about the efficacy of treatment to be made earlier. They can also be used to test subsequent entry complex drugs or agents with the same mechanism of action as an approved agent and support approval without a Phase 3 trial. They can also be used to demonstrate the efficacy of therapy in a pediatric population where this has already been demonstrated using intermediate and clinical outcomes in an adult MS population, thereby accelerating trials in the smaller, more vulnerable pediatric population.[69]

The use of intermediate MRI endpoints, such as brain atrophy or new/enlarging T2 lesions, has been effective in assessing DMTs that target inflammation, predominantly in individuals with RRMS. These endpoints have been useful because the effect of DMT on MRI lesions predicts the effect on relapses;[70] the association between MRI endpoints and worsening disability (EDSS) is weaker.[71] As therapies targeting pathobiologies other than inflammation emerge,[41] alternative intermediates will be needed.[41,72–74] To be useful, a clear understanding of the biological mechanism that the intermediate outcome assesses is needed, including how specific the outcome is to that mechanism. For example, can the effects of acute inflammation and neurodegeneration be differentiated? Therefore, an understanding of how the intermediate outcome relates to the clinical outcomes of interest for a Phase 3 trial is vital. Many potential biomarkers are being proposed as intermediates; these need to be prioritized and gaps and challenges related to test–retest reliability, inter- and intra-rater reliability, sample size calculations, and general availability addressed.[52,75]

Recommendations for future MS clinical trial research

Based on meeting discussions, attendees made several recommendations to enhance efficiency of clinical trials.

1. Access to clinical trial data sets is important to support trial design, and further development of data repositories accessible to qualified investigators is important.
2. Sustained funding, collaboration, and broad stakeholder engagement are needed to support platform trials. Opportunities for multi-national platform trials should be explored.
3. Consider evaluating and establishing a pipeline for testing candidate therapies using futility designs, then moving candidates that did not meet the futility threshold directly into a platform trial (Supplemental Appendix Figure e1). This may be particularly valuable for progressive MS and therapies targeting neuroprotection and repair strategies.
4. For DMTs being tested in pediatric MS that are already known to be effective in adults with MS, use Bayesian designs to reduce required sample sizes and likelihood of non-informative results.Intermediate outcomes may also substitute for clinical outcomes in this population. Alternatively, adaptive designs can be used to broaden trial inclusion criteria to progressively include children and youth, and older adults, over the course of the trial after careful consideration of potential differences in pharmacodynamics and pharmacokinetics between children and adults with MS.
5. Develop adaptive designs to test dose and duration of therapy required for remyelination and neuroprotective therapies, as well as rehabilitation interventions, and to identify the optimal population for treatment response.
6. Investigate acceptability to people with MS of linkages of clinical trial data with external sources to study long-term outcomes and develop standard consent language to support such data linkages.
7. Further develop, improve, and sustain existing disease registries to support pragmatic trials that can address questions such as comparative effectiveness of specific therapies or strategies or effects of specific health policies.
8. Develop biomarkers to enrich progressive MS trial populations with individuals who are likely to progress during the study period.
9. Prioritize intermediate outcomes for further development that target novel therapeutic mechanisms of action. Consideration should be given to cost, accessibility, reproducibility, and acceptability to patients:
   (a) Standardize terminology for intermediate and clinical outcomes and standardize methods of data collection.
   (b) Establish validity, reliability, responsiveness, and specificity of the intermediate outcome for the pathophysiological mechanism of interest.
   (c) Estimate required sample sizes.
   (d) Establish associations between the intermediate outcomes and meaningful clinical outcomes, including progression of physical and cognitive impairment and participation outcomes. This will require longitudinal studies—either new studies or augmenting data collection for new candidate intermediate outcomes in existing cohorts, and for a broader range of clinical outcomes that are meaningful to people with MS.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: R.A.M. received research funding from Canadian Institutes of Health Research, Research Manitoba, MS Canada, Multiple Sclerosis Scientific Foundation, Crohn's and Colitis Canada, National Multiple Sclerosis Society, Consortium of MS Centers, the Arthritis Society, and the US Department of Defense. She is supported by the Waugh Family Chair in Multiple Sclerosis. She is a co-investigator on a study funded in part by Biogen Idec and Roche (no funds to her or her institution). M.P.S. has received consulting fees from Biogen, Genzyme, GeNeuro, MedDay, Merck, Novartis, Roche, and Teva. F.B. received research funding from National Multiple Sclerosis Society and consulting fees from Biogen, Eisai, and Chiesi. W.Y.C. research funding from AstraZeneca, Janssen, and Roche. G.R.C. serves on Data and Safety Monitoring Boards for Applied Therapeutics, AI therapeutics, AMO Pharma, AstraZeneca, Avexis Pharmaceuticals, BioLineRx, Brainstorm Cell Therapeutics, Bristol Meyers Squibb/Celgene, CSL Behring, Galmed Pharmaceuticals, Green Valley Pharma, Horizon Pharmaceuticals, Immunic, Karuna Therapeutics, Mapi Pharmaceuticals Ltd, Merck, Mitsubishi Tanabe Pharma Holdings, Opko Biologics, Prothena Biosciences, Novartis, Regeneron, Sanofi-Aventis, Reata Pharmaceuticals, Teva Pharmaceuticals, NHLBI (Protocol Review Committee), University of Texas Southwestern, University of Pennsylvania, Visioneering Technologies, Inc., and consulting or advisory boards for Alexion, Antisense Therapeutics, Biogen, Clinical Trial Solutions LLC, Entelexo Biotherapeutics, Inc., Genzyme, Genentech, GW Pharmaceuticals, Immunic, Immunosis Pty Ltd, Klein-Buendel Incorporated, Merck/Serono, Novartis,

Perception Neurosciences, Protalix Biotherapeutics, Regeneron, Roche, and SAB Biotherapeutics. He is employed by the University of Alabama at Birmingham and President of Pythagoras, Inc. a private consulting company located in Birmingham AL. MW.K. received consulting fees and travel support from Biogen, Novartis, Roche, Sanofi Genzyme and EMD Serono. M.M.C. has no disclosures. E.M.M. has research/grant support from PCORI, National MS Society, National Institutes of Health, the US Department of Defense, Biogen, Teva, and Genentech. She received royalties for editorial duties from UpToDate and consulting fees from BeCareLink, LLC. J.J.H.P. is the founder of Core Clinical Sciences. F.P. has received research grants from Janssen, Merck KGaA and UCB, and fees for serving on DMC in clinical trials with Chugai, Lundbeck and Roche, and preparation of expert witness statement for Novartis. A.S. received research funding from MS Canada, National Multiple Sclerosis Society, CMSC and the US Department of Defence and is a member of editorial board for Neurology. She serves as a consultant for Gryphon Bio, LLC. She is a member of the Data and Safety Monitoring Board for Premature Infants Receiving Milking or Delayed Cord Clamping (PREMOD2), Central Vein Sign: A Diagnostic Biomarker in Multiple Sclerosis (CAVS-MS), and Ocrelizumab for Preventing Clinical Multiple Sclerosis in Individuals With Radiologically Isolated Disease (CELLO). In the last 3 years, J.C. has received support from the Efficacy and Evaluation (EME) Programme, a Medical Research Council (MRC) and National Institute for Health Research (NIHR) partnership and the Health Technology Assessment (HTA) Programme (NIHR), the UK MS Society, the US National MS Society, and the Rosetrees Trust. He is supported in part by the NIHR University College London Hospitals (UCLH) Biomedical Research Centre, London, UK. He has been a local principal investigator for a trial in MS funded by MS Canada. A local principal investigator for commercial trials funded by Ionis, Novartis, and Roche and has taken part in advisory boards/consultancy for Azadyne, Biogen, Lucid, Janssen, Merck, NervGen, Novartis, and Roche.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The International Advisory Committee on Clinical Trials in Multiple Sclerosis and the International Conference on Innovations in Clinical Trial Design and Enhancing Inclusivity of Clinical Trial Populations were supported by the National Multiple Sclerosis Society and the European Committee for Treatment and Research in Multiple Sclerosis. There was no involvement of the sponsors in the design, collection, analysis, or interpretation of data discussed at the conference. The opinions expressed are those of the authors.

ORCID iDs

Ruth Ann Marrie https://orcid.org/0000-0002-1855-5595
Maria Pia Sormani https://orcid.org/0000-0001-6892-104X
Gary R Cutter https://orcid.org/0000-0002-8455-980X
Marcus Werner Koch https://orcid.org/0000-0001-9972-5092
Morgan McCreary https://orcid.org/0000-0002-8378-460X
Fredrik Piehl https://orcid.org/0000-0001-8329-5219
Amber Salter https://orcid.org/0000-0002-1088-110X

Supplemental Material

Supplemental material for this article is available online.

References

1. Amatya B, Khan F and Galea M. Rehabilitation for people with multiple sclerosis: An overview of Cochrane Reviews. *Cochrane Database Syst Rev* 2019; 1: CD012732.
2. Feinstein A, Freeman J and Lo AC. Treatment of progressive multiple sclerosis: What works, what does not, and what is needed. *Lancet Neurol* 2015; 14: 194–207.
3. Cohen JA, Trojano M, Mowry EM, et al. Leveraging real-world data to investigate multiple sclerosis disease behavior, prognosis, and treatment. *Mult Scler* 2020; 26(1): 23–37.
4. Berger ML, Mamdani M, Atkins D, et al. Good research practices for comparative effectiveness research: Defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: The ISPOR good research practices for retrospective database analysis task force report—Part I. *Value Health* 2009; 12(8): 1044–1052.
5. Loudon K, Treweek S, Sullivan F, et al. The PRECIS-2 tool: Designing trials that are fit for purpose. *BMJ* 2015; 350: h2147.
6. Li G, Sajobi TT, Menon BK, et al. Registry-based randomized controlled trials—What are the

advantages, challenges, and areas for future research? *J Clin Epidemiol* 2016; 80: 16–24.

7. U.S. Food and Drug Administration. Evaluating inclusion and exclusion criteria in clinical trials, https://www.fda.gov/media/134754/download
8. U.S. Department of Health and Human Services. Benefit-risk assessment for new drug and biological products: Guidance for industry DRAFT GUIDANCE, https://www.regulations.gov/document/FDA-2020-D-2316-0002
9. Ontaneda D, Tallantyre EC, Raza PC, et al. Determining the effectiveness of early intensive versus escalation approaches for the treatment of relapsing-remitting multiple sclerosis: The DELIVER-MS study protocol. *Contemp Clin Trials* 2020; 95: 106009.
10. Johns Hopkins University. *Traditional versus early aggressive therapy for multiple sclerosis trial (TREAT-MS, NCT03500328)*. Washington, DC: National Library of Medicine, 2018.
11. Lauer MS and D'Agostino RB. The randomized registry trial—The next disruptive technology in clinical research? *New Engl J Med* 2013; 369: 1579–1581.
12. Song L, Hu X, Ma L, et al. INTEnsive care bundle with blood pressure reduction in acute cerebral hemorrhage trial (INTERACT3): Study protocol for a pragmatic stepped-wedge cluster-randomized controlled trial. *Trials* 2021; 22: 943.
13. Li F and Wang R. Stepped wedge cluster randomized trials: A methodological overview. *World Neurosurg* 2022; 161: 323–330.
14. Dimairo M, Pallmann P, Wason J, et al. The Adaptive designs CONSORT Extension (ACE) statement: A checklist with explanation and elaboration guideline for reporting randomised trials that use an adaptive design. *BMJ* 2020; 369: m115.
15. Selmaj K, Li DKB, Hartung HP, et al. Siponimod for patients with relapsing-remitting multiple sclerosis (BOLD): An adaptive, dose-ranging, randomised, phase 2 study. *Lancet Neurol* 2013; 12(8): 756–767.
16. Wason JMS, Brocklehurst P and Yap C. When to keep it simple—Adaptive designs are not always useful. *BMC Med* 2019; 17: 152.
17. Park JJH, Siden E, Zoratti MJ, et al. Systematic review of basket trials, umbrella trials, and platform trials: A landscape analysis of master protocols. *Trials* 2019; 20: 572.
18. Park JJH, Detry MA, Murthy S, et al. How to use and interpret the results of a platform trial: Users' guide to the medical literature. *JAMA* 2022; 327: 67–74.
19. Renfro LA and Sargent DJ. Statistical controversies in clinical research: Basket trials, umbrella trials, and other master protocols: A review and examples. *Ann Oncol* 2017; 28: 34–43.
20. Parker CC, James ND, Brawley CD, et al. Radiotherapy to the primary tumour for newly diagnosed, metastatic prostate cancer (STAMPEDE): A randomised controlled phase 3 trial. *Lancet* 2018; 392: 2353–2366.
21. Carthon BC and Antonarakis ES. The STAMPEDE trial: Paradigm-changing data through innovative trial design. *Transl Cancer Res* 2016; 5(Suppl. 3): S485–S490.
22. Park JJH, Sharif B, Harari O, et al. Economic evaluation of cost and time required for a platform trial vs conventional trials. *JAMA Netw Open* 2022; 5: e2221140.
23. Schiavone F, Bathia R, Letchemanan K, et al. This is a platform alteration: A trial management perspective on the operational aspects of adaptive and platform and umbrella protocols. *Trials* 2019; 20: 264.
24. Kumar-Sinha C and Chinnaiyan AM. Precision oncology in the age of integrative genomics. *Nat Biotechnol* 2018; 36: 46–60.
25. Koch MW, Korngut L, Patry DG, et al. The promise of futility trials in neurological diseases. *Nat Rev Neurol* 2015; 11(5): 300–305.
26. Koch MW, Kaur S, Sage K, et al. Hydroxychloroquine for primary progressive multiple sclerosis. *Ann Neurol* 2021; 90: 940–948.
27. Gehan EA. The determination of the number of patients required in a preliminary and a follow-up trial of a new chemotherapeutic agent. *J Chron Dis* 1961; 13: 346–353.
28. Fleming TR. One-sample multiple testing procedure for phase II clinical trials. *Biometrics* 1982; 38(1): 143–151.
29. Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials* 1989; 10: 1–10.
30. Koch MW, Sage K, Kaur S, et al. Repurposing domperidone in secondary progressive multiple sclerosis: A Simon 2-stage phase 2 futility trial. *Neurology* 2021; 96: e2313–e2322.
31. Fors M and González P. Current status of Bayesian clinical trials for oncology, 2020. *Contemp Clin Trials Commun* 2020; 20: 100658.
32. Bonangelino P, Irony T, Liang S, et al. Bayesian approaches in medical device clinical trials: A discussion with examples in the regulatory setting. *J Biopharm Stat* 2011; 21(5): 938–953.
33. Tan SB, Dear KB, Bruzzi P, et al. Strategy for randomised clinical trials in rare cancers. *BMJ* 2003; 327: 47–49.
34. Schmidli H, Gsteiger S, Roychoudhury S, et al. Robust meta-analytic-predictive priors in clinical

trials with historical control information. *Biometrics* 2014; 70(4): 1023–1032.

35. Zhang X, Cutter G and Belin T. Bayesian sample size determination under hypothesis tests. *Contemp Clin Trials* 2011; 32(3): 393–398.
36. Berry DA. Bayesian clinical trials. *Nat Rev Drug Discov* 2006; 5: 27–36.
37. Sormani MP and Waubant E. Paediatric multiple sclerosis: A lesson from TERIKIDS. *Lancet Neurol* 2021; 20(12): 971–973.
38. Bovis F, Ponzano M, Signori A, et al. Reinterpreting clinical trials in children with multiple sclerosis using a Bayesian approach. *JAMA Neurol* 2022; 79: 821–822.
39. Cohen JA, Calabresi PA, Chakraborty S, et al. Avonex combination trial in relapsing-remitting MS: Rationale, design and baseline data. *Mult Scler* 2008; 14(3): 370–382.
40. Conway D and Cohen JA. Combination therapy in multiple sclerosis. *Lancet Neurol* 2010; 9: 299–308.
41. Kuhlmann T, Moccia M, Coetzee T, et al. Multiple sclerosis progression: Time for a new mechanism-driven framework. *Lancet Neurol* 2023; 22(1): 78–88.
42. Yuan Y, Lee JJ and Hilsenbeck SG. Model-assisted designs for early-phase clinical trials: Simplicity meets superiority. *JCO Precis Oncol* 2019; 3: 1–12.
43. Anderson RL, DiMeglio LA, Mander AP, et al. Innovative designs and logistical considerations for expedited clinical development of combination disease-modifying treatments for type 1 diabetes. *Diabet Care* 2022; 45: 2189–2201.
44. Feinstein A, Amato MP, Brichetto G, et al. Study protocol: Improving cognition in people with progressive multiple sclerosis: A multi-arm, randomized, blinded, sham-controlled trial of cognitive rehabilitation and aerobic exercise (COGEx). *BMC Neurol* 2020; 20: 204.
45. Fitzpatrick T, Perrier L, Shakik S, et al. Assessment of long-term follow-up of randomized trial participants by linkage to routinely collected data: A scoping review and analysis. *JAMA Netw Open* 2018; 1: e186019.
46. The IFNB Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group. Interferon beta-lb in the treatment of multiple sclerosis: Final outcome of the randomized controlled trial. *Neurology* 1995; 45: 1277–1285.
47. Goodin DS, Reder AT, Ebers GC, et al. Survival in MS: A randomized cohort study 21 years after the start of the pivotal IFNbeta-1b trial. *Neurology* 2012; 78: 1315–1322.
48. Aasdahl L, Vasseljen O, Gismervik SØ, et al. Two-year follow-up of a randomized clinical trial of inpatient multimodal occupational rehabilitation vs outpatient acceptance and commitment therapy for sick listed workers with musculoskeletal or common mental disorders. *J Occup Rehabil* 2021; 31: 721–728.
49. Burcu M, Manzano-Salgado CB, Butler AM, et al. A framework for extension studies using real-world data to examine long-term safety and effectiveness. *Ther Innov Regul Sci* 2022; 56(1): 15–22.
50. Hay AE, Leung YW, Pater JL, et al. Linkage of clinical trial and administrative data: A survey of cancer patient preferences. *Curr Oncol* 2017; 24(3): 161–167.
51. Hemming K, Hutton JL, Maguire MJ, et al. Open label extension studies and patient selection biases. *J Eval Clin Pract* 2008; 14: 141–144.
52. Seeger JD, Davis KJ, Iannacone MR, et al. Methods for external control groups for single arm trials or long-term uncontrolled extensions to randomized clinical trials. *Pharmacoepidemiol Drug Saf* 2020; 29: 1382–1392.
53. Bohensky MA, Jolley D, Sundararajan V, et al. Development and validation of reporting guidelines for studies involving data linkage. *Aust N Z J Public Health*. 2011; 35: 486–489.
54. McAdams M, Stankiewicz JM, Weiner HL, et al. Review of phase III clinical trials outcomes in patients with secondary progressive multiple sclerosis. *Mult Scler Relat Disord* 2021; 54: 103086.
55. Nowinski CJ, Miller DM and Cella D. Evolution of patient-reported outcomes and their role in multiple sclerosis clinical trials. *Neurotherapeutics* 2017; 14(4): 934–944.
56. Kuspinar A and Mayo NE. A review of the psychometric properties of generic utility measures in multiple sclerosis. *Pharmacoeconomics* 2014; 32(8): 759–773.
57. Rotstein D, Solomon JM, Sormani MP, et al. Association of no evidence of disease activity with no long-term disability progression in multiple sclerosis: A systematic review and meta-analysis. *Neurology* 2022; 99: e209–e220.
58. Motl RW, Cohen JA, Benedict R, et al. Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. *Mult Scler* 2017; 23(5): 704–710.
59. Feys P, Lamers I, Francis G, et al. The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. *Mult Scler* 2017; 23(5): 711–720.
60. Benedict RH, DeLuca J, Phillips G, et al. Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. *Mult Scler* 2017; 23(5): 721–733.

61. Song MK, Lin FC, Ward SE, et al. Composite variables: When and how. *Nurs Res* 2013; 62: 45–49.
62. Fischer JS, Jak AJ, Kniker JE, et al. *Administration and scoring manual for the Multiple Sclerosis Functional Composite measure (MSFC)*. New York: Demos Medical Publishing, 2001.
63. Freemantle N, Calvert M, Wood J, et al. Composite outcomes in randomized trials: Greater precision but with greater uncertainty? *JAMA* 2003; 289: 2554–2559.
64. Kapoor R, Ho PR, Campbell N, et al. Effect of natalizumab on disease progression in secondary progressive multiple sclerosis (ASCEND): A phase 3, randomised, double-blind, placebo-controlled trial with an open-label extension. *Lancet Neurol* 2018; 17(5): 405–415.
65. DeMets DL, Psaty BM and Fleming TR. When can intermediate outcomes be used as surrogate outcomes? *JAMA* 2020; 323: 1184–1185.
66. Holper L, Coenen M, Weise A, et al. Characterization of functioning in multiple sclerosis using the ICF. *J Neurol* 2010; 257(1): 103–113.
67. Finlayson M, Feys P, Dalgas U, et al. Intermediate outcomes in MS rehabilitation: Needs, challenges and future directions to strengthen clinical trials, *Mult Scler* 2023 (Submitted).
68. Korn EL and Freidlin B. Surrogate and intermediate endpoints in randomized trials: What's the goal? *Clin Canc Res* 2018; 24: 2239–2240.
69. Abdel-Mannan O, Ciccarelli O and Hacohen Y. Considering the future of pediatric multiple sclerosis trials after the CONNECT open-label randomized trial. *JAMA Netw Open* 2022; 5: e2230451.
70. Sormani MP and Bruzzi P. MRI lesions as a surrogate for relapses in multiple sclerosis: A meta-analysis of randomised trials. *Lancet Neurol* 2013; 12(7): 669–676.
71. Sormani MP, Bonzano L, Roccatagliata L, et al. Surrogate endpoints for EDSS worsening in multiple sclerosis. A meta-analytic approach. *Neurology* 2010; 75: 302–309.
72. Granziera C, Wuerfel J, Barkhof F, et al. Quantitative magnetic resonance imaging towards clinical application in multiple sclerosis. *Brain* 2021; 144: 1296–1311.
73. Oh J, Ontaneda D, Azevedo C, et al. Imaging outcome measures of neuroprotection and repair in MS: A consensus statement from NAIMS. *Neurology* 2019; 92: 519–533.
74. Dangond F, Donnelly A, Hohlfeld R, et al. Facing the urgency of therapies for progressive MS—A Progressive MS Alliance proposal. *Nat Rev Neurol* 2021; 17(3): 185–192.
75. Shinohara RT, Oh J, Nair G, et al. Volumetric analysis from a harmonized multisite brain MRI study of a single subject with multiple sclerosis. *AJNR Am J Neuroradiol* 2017; 38(8): 1501–1509.